COMPUTATIONAL SENSING AND IN VITRO CLASSIFICATION OF GMOs AND BIOMOLECULAR EVENTS

Elebeoba May[1]*, Miler T. Lee[2]†, Patricia Dolan[1], Paul Crozier[1], Susan Brozik[1], Monica Manginell[1]

[1]Sandia National Laboratories, Albuquerque, NM USA ‡
[2]University of Pennsylvania School of Medicine, Philadelphia, PA USA

ABSTRACT

The increasing threat of microbial weapons of mass destruction (WMD) creates a critical need for rapid development and deployment of target-specific microsensor detection systems. Next-generation biosensor-dependent technologies must incorporate intelligence into the microsensor platform, enabling the execution of complex detection tasks, such as systematic identification and classification of microbial agents and genetically modified organisms (GMOs) in the presence of non-lethal agents. Using an information and coding-theoretic framework we develop a *de novo* method for mapping a generic bio-detection and classification problem to deoxyribozyme-compliant algorithms. We implement an algorithm on our novel computational biosensor system and as a proof-of-concept develop an application-specific, computational biosensor for concurrent detection and classification of H5N1 regional strains.

1. INTRODUCTION

Hybridization-based target recognition and discrimination is central to a wide variety of applications: high throughput screening, distinguishing genetically modified organisms (GMOs), molecular computing, differentiating biological markers, fingerprinting a specific sensor response for complex systems, etc. The recognition substrate can exist in solution or be immobilized onto a transducer and hybridization events can be detected optically, electrochemically, or via a mass-sensitive device [11]. The bioreceptor, or probe, is critical to the specificity of the biosensor. Although several single-stranded DNA sensor technologies, such as DNA microarrays, are widely used, molecular beacon probes are highly sensitive and specific bioreceptors [11, 3, 12]. They can detect mutations in target sequences and can be multiplexed [3]; these properties make molecular beacons effective platforms for detecting genetically modified targets in biodefense systems.

1.1 Deoxyribozyme Molecular Beacons

Molecular beacons are single-stranded oligonucleotide probes that form stem-loop structures. The loop contains a probe sequence that is complementary to a target sequence. Traditional molecular beacons contain a fluorophore and quencher on each arm of the stem of the beacon. Fluorescence is achieved by separation of the fluorophore and quencher due to a conformation change that takes place following target hybridization to the loop structure [9]. However, traditional beacons are limited by a 1:1 (target: signal) stoichiometry, and the sensitivity of the detection is linked to the amount of target present. Target amplification (PCR) is required to increase the level of sensitivity.

Though DNA serves primarily as a carrier of the genetic code and no enzymes made of DNA have been found in nature, molecular beacons comprised of single-stranded DNAs can be engineered to perform catalytic reactions similar to those of protein and RNA. Catalytic DNAs or deoxyribozymes are synthesized in the laboratory via an in vitro iterative selection process.

Like traditional molecular beacons, catalytic molecular beacons also contain a stem-loop structure that undergoes a conformational change following target hybridization to the loop region. Unlike traditional molecular beacons, catalytic molecular beacons are modular molecules, containing a deoxyribozyme appended to the stem-loop structure [9]. Catalytic activity is initiated by a target DNA sequence binding to the loop region that is distinct from the enzymatic active site. This allosterically activates the deoxyribozyme complex to bind and cleave a labeled substrate oligonucleotide molecule, producing a detectable signal. Once activated, the catalytic molecular beacon will continuously cleave labeled substrate molecules. Target amplification is not required since signal amplification is obtained through repeated pro-

*Corresponding author: eemay@sandia.gov
†M. Lee is a U.S. Department of Energy Computational Sciences Graduate Fellow in the Genomics and Computational Biology Program.
‡E. May is with the Computational Biology Department; P. Dolan, S. Brozik, and M. Manginell are with the Biosensors and Nanomaterials Department; P. Crozier is with the Multiscale Computational Materials and Methods Department. Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company for the United States Department of Energy's National Nuclear Security Administration under contract DEAC0494AL85000.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
DEC 2008	**N/A**	**-**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Computational Sensing And In Vitro Classification Of Gmos And Iomolecular Events	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Sandia National Laboratories, Albuquerque, NM USA**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES
See also ADM002187. Proceedings of the Army Science Conference (26th) Held in Orlando, Florida on 1-4 December 2008, The original document contains color images.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**UU**	**4**	

cessing of excess substrate molecules due to the recognition and binding of a single target. Single and multi-receptor site configurations enable the deoxyribozyme to function as YES, NOT, or multi-input AND gates [10].

Current high-throughput DNA sensing systems rely heavily on silicon-based computing for interpretation of molecular recognition events. Complex bioinformatics algorithms are tasked with de-noising and processing of sensor output signals. This approach is error prone and not easily integrated into emerging nanoscale sensor architectures for lab-on-chip systems. Molecular computation, although currently impractical for real time computing needs, offers the ability for massively parallel computation and detection, and is a promising technology base for the development of next generation hybrid molecular sensor systems. *In vitro* computation using deoxyribozyme gates enables the design of complex biosensor systems and reduces reliance on classical computing for interpretation of sensor response.

In this work we describe the development of a deoxyribozyme-based computational biosensor. In Section 2 of this paper we discuss the methods for developing a computational biosensor for concurrent detection and classification of targets using deoxyribozyme gates. In Section 3 we demonstrate the use of our sensor platform in the de novo design and development of biosensor systems for detecting genetic modifications in avian influenza and identifying a sample strain's region of origin.

2. METHODS

We have devised a multi-step process for rapid *de novo* design, development and fabrication of target-specific microsensor detection systems. Given a "target," we begin with a bioinformatical analysis and identify an initial set of candidate probe sequences. Next, using the initial probe set, we generate a problem specific detection/classification algorithm. The type and complexity of the algorithm depends on the problem our system will evaluate. Phase two produces multiple intermediate probe sets with corresponding deoxyribozyme-compliant algorithms. This information is the input to phase three, during which we evaluate the feasibility of the candidate probe sets and associated algorithms. Gate- and system-level performance is verified through simulation. The optimal design then enters phase four, which consists of system fabrication and experimental validation. We apply our process to the *de novo* design and development of computational biosensors for 1) detection of key mutations in avian influenza as reported in [5]; 2) detection of geographical strains of avian influenza, which is described in this work.

2.1 Distinguishing Geographical Strains of H5N1

Surveillance plus antigenic and genetic analysis are used to monitor known H5N1 variants and identify emerging sublineages [8]. Continued monitoring in regional hotspots, such as southern China and Southeast Asia, can aid in evaluating the effectiveness of disease control protocols and in identifying the emergence of potentially pandemic strains. Large-scale sequencing efforts, such as the St. Jude Influenza Genome Project [7], provide genetic information for the design of computational biosensors that can aid in tracking and characterizing regional strains of H5N1. We discuss the design and feasibility of such systems.

2.2 Informational analysis and extraction of candidate marker regions.

We downloaded hemagglutinin (HA) gene sequences of H5N1 influenza samples from the NCBI Influenza Virus Resource [1]. In all, we used 164 sequences from samples spanning five different geographic locales: Egypt, Hong Kong, Indonesia, Thailand, and Vietnam. We generated a multiple alignment and consensus sequence for each locale using MUSCLE [2] with the default parameters, then created a multiple alignment of the five consensus sequences, in order to index the nucleotides across all sequences uniformly. To find regions of low variation within a locale, we used an information theory approach that yielded 996 candidate probes. We reduced the region classification problem to an experimentally tractable form: develop a deoxyribozyme-compliant algorithm that detects and classifies a candidate strain as originating from Egypt, Hong Kong, or Indonesia, using a reduced probe set of 75 candidate probes. The reduced probe set was selected from the 996 candidate set based on their classification sensitivity values. We further pruned our probe set to 52 candidate probes by filtering out probes that did not produce viable deoxribozyme structures or produced undesirable secondary structures.

2.3 Design of realizable deoxyribozyme-compliant classification algorithms.

Using methods from communication and coding theory [6, 4] we generate constraints for producing an efficient deoxyribozyme-compliant classification algorithm using a minimal set of probes. We devised a reduced (N=52,K=47) coding-based algorithm for subtype classification that uses nine probes to detect and classify an H5N1 nucleic acid sample as originating from Egypt, Hong Kong, or Indonesia. The three-bit classification algorithm is: CompWell1 = probe23 $\oplus$ probe48; CompWell2 = probe4 $\oplus$ probe23 $\oplus$ probe24 $\oplus$ probe34 $\oplus$ probe43 $\oplus$ probe50; CompWell3 = probe20 $\oplus$ probe51. CompWell represents the computational well, which is realized by combining the discretized binary output of the microplate wells.

Prior to fabrication and experimental testing, we can computationally "test" the gate using hybridization thermodynamics to predict the fluorescence response of our gate to target sequences and possibly mutated forms of the target [5]. For the YES_{i34} gate, which would bind the complement of probe 34, we used a set of optimized fit parameters to predict the response of the gate to its complementary target sequence and fifteen mutant inputs (Figure 1). For the region identification problem, we desire gates that can accurately distinguish the target from a non-target; the gates do not need to exhibit a distinct, mutation-dependent fluorescence signature. Based on our initial predictive simulation of the YES_{i34} gate, it seems to be a viable gate and would potentially pass on to the fabrication process. We can perform similar tests on each generated gate.

We used the Oligonucleotide Modeling Platform (OMP) from DNA Software to compute parameters to predict the response of individual deoxyribozyme gates to input probes and to simulate the algorithmic performance of the sys-

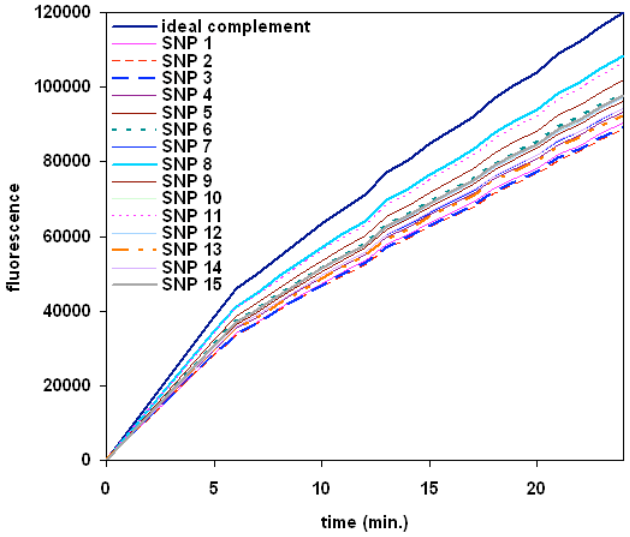

Figure 1: Predicted ribozyme fluorescence output for the designed YES_{i34} gate (5'-CTCGTCTGTTCCGAT-3') and corresponding full complement plus 15 SNP-containing input target sequences. Fluorescence values given in arbitrary units.

tem. Modeling and simulation enables us to detect undesirable hybridizations that may occur in the presence of input probes or when combined to form computational wells. Upon detection we can return to our algorithm generation step and produce a more realizable algorithm. This reduces the probability of fabricating faulty systems, thus reducing final cost.

2.4 Realization of Computational Biosensor

We implement a single-gate (YES gate) only system and a multi-gate (YES-gate/AND-NOT gate) system. Figure 2 shows single and double input deoxyribozyme gates. Single

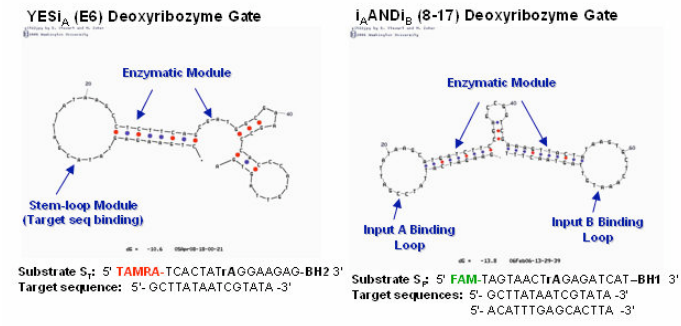

Figure 2: YES_{iA} (E6) and $_{iA}AND_{iB}$ (8-17) Gates.

gates use only one probe, while multi-gates are composed of two or three probes. The AND-NOT gate implementation uses two AND-NOT gates to implement the exclusive-OR

Table 1: Probe composition of "sample" test sequences (HK=Hong Kong, Indo=Indonesia) .

Seq	P_4	P_{20}	P_{23}	P_{24}	P_{34}	P_{43}	P_{48}	P_{50}	P_{51}
Egypt	1	0	0	0	0	0	0	0	0
HK	0	0	0	0	0	1	1	1	1
Indo-1	0	1	1	1	1	0	0	0	0
Indo-2	0	1	0	0	1	0	0	0	0
Indo-3	0	1	1	1	0	0	0	0	0

($\oplus$) function. Currently we perform the exclusive-OR in our post-processing step. But the AND-NOT gate demonstrates our ability to perform the exclusive-OR operation in a microplate well. Experimental tests were run to assess the computational biosensing platform:

2.5 Experimental Design

The following experimental tests were run to assess the intelligent biosensing platform:

Control 1: Test each of the 9 YES_{iA} gates to ensure operability and to produce a reference signal.

Control 2: Test each of the 4 AND-NOT gates to ensure operability and produce reference signal.

Control 3: Test each XOR (consists of two AND-NOT gates in a single well

$$probe_1 \oplus probe_2 \quad = \quad {}_{probe1}AND - NOT_{probe2} + {}_{probe1}AND - NOT_{probe2}$$

where + represents OR function.

- Test1: Probe1 absent, Probe2 absent
- Test2: Probe1 absent, Probe2 present
- Test3: Probe1 present, Probe2 absent
- Test4: Probe1 present, Probe2 present

Classify: For the regional subtype classification experiments, we add a mixture containing all the probes indicated by a 1 (the probes are subsequence found in the sample sequences downloaded from NCBIs Influenza Virus Resource [1]) into each of the microplate wells (subwells) to simulate the test "sample". Table 1 below shows the "samples" and the probes (P_i) present in the sample.

3. EXPERIMENTAL RESULTS

3.1 Feasibility demonstration using a flourescence-based laboratory platform.

We devised a fluorescence-based laboratory detection platform, designed, and constructed all required deoxyribozyme gates to implement a rapid diagnostic system for subtype classification for avian influenza strain H5N1. Our experimental setup consists of a 55-60?l detection volume containing: reaction buffer, deoxyribozyme gates, fluorophore-labeled substrates, and input sequences. Fluorescence was measured with a FluoDiaT70 Microplate Reader (PTI, Inc.) using black 384-well microplates (OptiPlate-384F, Perkin-Elmer) at room temperature.

Table 2: Syndrome classification patterns for subtype classification.

Subtype Class	CompWell_1	CompWell_2	CompWell_3
Egypt	0	1	0
Hong Kong	1	0	1
Indonesia	1	1	1

3.2 Classification via computational post-processing

In comparison to traditional hybridization-based sensors, the output of the computational biosensor requires limited post-processing to identify the subtype category to which the sample nucleic acid sequence belongs. We first discretize the fluorescence output from each microplate well (subwell) to a binary value by comparing the slope of the linear regression line of the control to that of the sample. If the slope of the sample is equal to or greater than $\frac{1}{3}$ of the slope of the control, that subwell is scored as a binary 1, else it is scored as a binary 0. The binary subwell values are used to compute the value of each of the three CompWells. The pattern of the computational wells determines what group the sequence belongs. Table 2 shows the expected binary patterns for each group. Using the binary patterns, we correctly identified the subtype of 97.3% of the "sample" sequences.

4. CONCLUSION

We have developed a platform for the de novo design and realization of a computational biosensor based on deoxyribozyme logic gates. The biosensor system is able to detect genetic modifications in the target and perform complex detection tasks, reducing reliance on classical computing for output interpretation. The ability to compute differentiates this system from other molecular beacon based systems. By increasing the complexity of the algorithm (or incorporating more computational wells) we can potentially increase the sensitivity and specificity of the biosensor, hence increasing accuracy by increasing computation.

Our deoxyribozyme-based computational sensor system can provide rapid detection (in optical detection we were able to distinguish target and mutated target within minutes), will not require PCR for target amplification, can simultaneously detect and classify in vitro, and is favorable for a field deployable platform. This *de novo* platform provides a unique capability that can help transform the way we detect biological agents and protect the Future Soldier against microbial WMDs.

We are investigating immobilization methods for electrochemical detection of deoxyribozyme gate activity. The electrochemical detection platform will produce differential currents to indicate positive or negative algorithm outcomes. This will obviate our need for a fluorescence reader, thus permitting the development of a field deployable rapid detection system

Acknowledgment

The authors would like to thank M. Finley (Sandia National Laboratories) for insights on avian influenza surveillance; M. Stojanovic (Columbia University), D. Stefanovic (University of New Mexico), and J. Macdonald (Columbia University) for insights on deoxyribozyme gates. M. Lee was supported by the U.S. DOE CSGF #DE-FG02-97ER25308. This project was funded by Sandia National Laboratories' Laboratory Directed Research and Development program.

5. REFERENCES

[1] Influenza virus resource, 2004. [cited 2007 Aug 8].

[2] R. C. Edgar. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*, 32(5):1792–7, 2004.

[3] D. Horejsh, F. Martini, F. Poccia, G. Ippolito, A. DiCaro, and M. Capobianchi. A molecular beacon, bead-based assay for the detection of nucleic acids by flow cytometry. *Nucleic Acids Research*, 33(2), 2005.

[4] S. Lin and D. J. Costello. *Error Control Coding: 2nd Edition*. Prentice-Hall, Inc., Englewood Cliffs, NJ, 2004.

[5] E. May, P. Dolan, P. Crozier, S. Brozik, and M. Manginell. Towards De novo design of Deoxyribozyme Biosensors for GMO detection. *IEEE Sensors Journal,*, June 2008.

[6] E. E. May, M. A. Vouk, and D. L. Bitzer. An Error-Control Coding Model For Classification of Escherichia coli K-12 Ribosome Binding Sites. *IEEE EMB Magazine*, January 2006.

[7] J. Obenauer, J. Denson, P. Mehta, X. Su, S. Mukatira, D. B. Finkelstein, X. Xu, J. Wang, J. Ma, Y. Fan, K. Rakestraw, R. Webster, E. Hoffmann, S. Krauss, J. Zheng, Z. Zhang, and C. Naeve. Large-scale sequence analysis of avian influenza isolates. *Science*, 311:1576–1580, March 2006.

[8] G. Smith, X. Fan, J. Wang, K. Li, K. Qin, J. Zhang, D. Vijaykrishna, C. Cheung, K. Huang, J. Rayner, J. Peiris, H. Chen, R. Webster, and Y. Guan. Emergence and predominance of an H5N1 influenza variant in China. *PNAS*, 103(45):16936–16941, November 2006.

[9] M. N. Stojanovic, P. de Prada, and D. W. Landry. Catalytic molecular beacons. *Chembiochem*, 2:411–415, 2001.

[10] M. N. Stojanovic and D. Stefanovic. A deoxyribozyme-based molecular automaton. *Nature Biotechnology*, 21:1069–107, 2003.

[11] J. Wang. Survey and summary: From DNA biosensors to gene chips. *Nucleic Acids Research*, 28(16):3011–3016, 2000.

[12] Z.-S. Wu, J.-H. Jian, G.-L. Shen, and R.-Q. Yu. Highly sensitive DNA detection and point mutation identification: an electrochemical approach based on the combined use of ligase and reverse molecular beacon. *Human Mutation*, 28(6):630–637, 2007.